Empowering Citizens. Smarter Societies.

Insight
Centre for Data Analytics

Analytics-driven Public Service and Policy Innovation

Adegboyega Ojo, Insight Centre for Data Analytics, Data Science Institute, NUI Galway, adegboyega.ojo@insight-centre.org

A World Leading SFI Research Centre

DCU NUI Galway OÉ Gaillimh UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh UCD Dublin Science Foundation Ireland sfi For what's next

Outline of talk

1) The changing government technology innovation environment

2) Data analytics in government

3) Data analytics for supporting services to protect vulnerable children

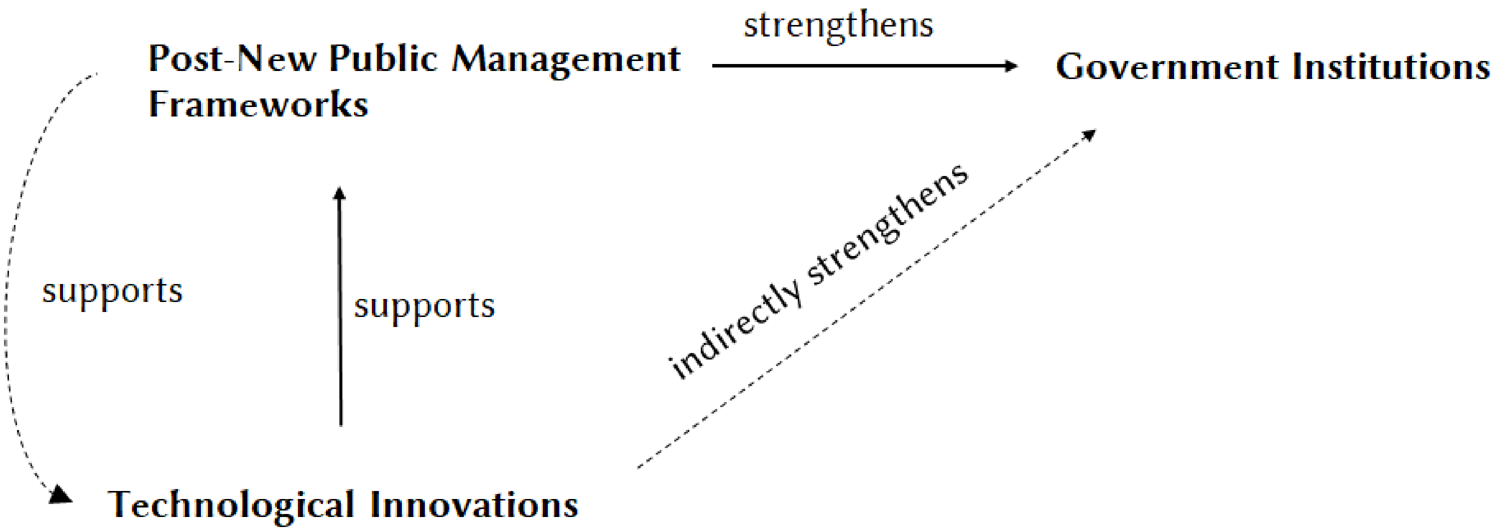
4) Excerpts of a data analytics in government project currently being implemented by Insight Centre (NUI Galway) in collaboration with government agencies

5) Challenges in developing data analytics capability in government

The new context for govtech innovation

Insight
Centre for Data Analytics

DEG PVM NPG



The duality of new public management & gov tech innovation

Ojo A. (2019) Next Generation Government - Hyperconnected, Smart and Augmented. In: Camarinha-Matos L., Afsarmanesh H., Antonelli D. (eds) Collaborative Networks and Digital Transformation. PRO-VE 2019. IFIP Advances in Information and Communication Technology, vol 568. Springer, Cham

The new context for govtech innovation

Insight
Centre for Data Analytics

Post-NPM Paradigm Affordances

1. Easy access to government information and data
2. Engaging citizens in co-production to address societal challenges
3. Provisioning shared and integrated services within a joined-up thinking framework
4. Focus public value creation
5. Understanding public and collective interests
6. Creating innovation-friendly environment within government
7. Public-Private partnership and collaboration networks

HYPEROPEN GOV CAPABILITY:

- Highly personalised information & knowledge delivery to citizen over traditional and new media
- Virtual Cognitive agents responding to knowledge queries

DIYGOV CAPABILITY

- Citizen/business-initiated co-production of digital services anytime
- Access to tools to rapidly develop new services
- Discovery and proactive notification of citizens on services of interest

HYPERCOLLABORATIVE GOV CAPABILITY

- Dynamic Collaborative Network of State & Non-State actors
- Automated discovery of potential network partners
- Automatic reconfiguration of collaboration network

Technological Innovation Affordances

1. Harness different forms of data on government operations, services & infrastructure.
2. Predicting individual information and public service needs.
3. Inter-organisational collaboration and information sharing
4. Tools for participation, co-production and bottom-up engagement
5. Task automation & cognitive engagement for question answering
6. Predictive & prescriptive analytics
7. Knowledge discovery and synthesis (induction)

Ojo A. (2019) Next Generation Government - Hyperconnected, Smart and Augmented. In: Camarinha-Matos L., Afsarmanesh H., Antonelli D. (eds) Collaborative Networks and Digital Transformation. PRO-VE 2019. IFIP Advances in Information and Communication Technology, vol 568. Springer, Cham

Data Analytics in the Public Sector

Insight
Centre for Data Analytics

- **Data analytics is the discovery, interpretation, and communication of meaningful patterns in data.**

- Analytics can used by both individual and multiple teams/ organisations to better inform their own decisions and activities and collaborate more effectively.

- The value of data analytics is associated with the improvement in outcomes of decisions or actions it affords

https://media.nesta.org.uk/documents/Public_Sector_Data_Analytics_-_A_Nesta_Guide_byCwKTI.pdf

Data Analytics in the Public Sector

Insight
Centre for Data Analytics

Data analytics can be used to support the following in government environment:

- Identifying specific cases in a wider group
- Prioritising cases based on risk or need
- Creating early warning tools
- Making better, quicker decisions • Optimising resource allocation

https://media.nesta.org.uk/documents/Public_Sector_Data_Analytics_-_A_Nesta_Guide_byCwKTI.pdf

The Analytics Imperative

Insight
Centre for Data Analytics

Elected officials and chiefs of staff must address analytics now!

"Analytics is much more than a new technology trend. It represents a paradigm shift, upending the way people think, plan and act and that includes those leading public service agencies. Because of its potential, government analytics puts a new and pressing responsibility squarely on the shoulders of public officials. Those who fail to adopt government analytics as policy are essentially up against a team they can't beat; tight budgets, pressing demands and citizens who are increasingly impatient for results!"

Moneyball Under the Dome - Government Analytics for Public Officials, Accenture, 2014

AI & Data Analytics in Public Sector

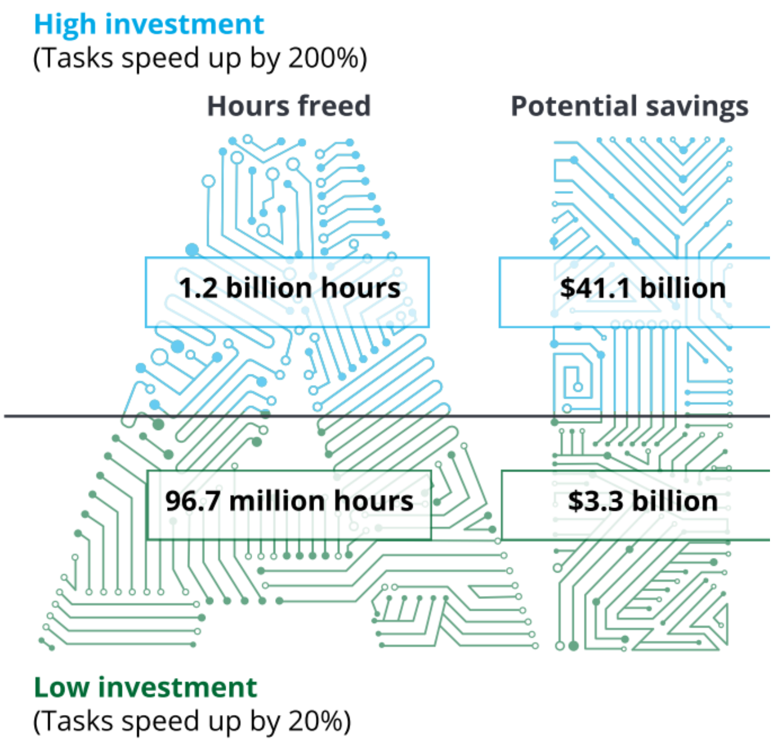
According to the European AI Strategy, AI will transform public services [1].

Capgemini estimates that AI applications in the public sector will create a savings of between €2 and 5 billion globally and 1.93% point growth in world GDP by 2025 [2].

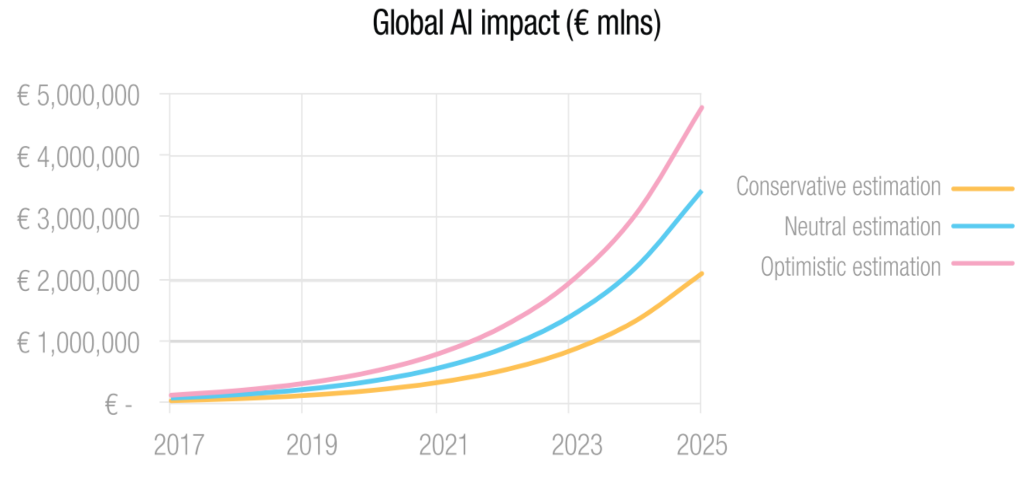
"*The most challenging problems AI may help us solve—from fighting terrorists to serving vulnerable populations—will involve government. More immediately, though not less consequentially, AI will change the way public servants do their jobs.*" [3]

1) https://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-795-F1-EN-MAIN-PART-1.PDF
2) https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/10/ai-in-public-sector.pdf
3) http://www.businessofgovernment.org/sites/default/files/Using%20Artificial%20Intelligence%20to%20Transform%20Government.pdf

AI & Data Analytics in the Public Sector



Source: Deloitte analysis.



Global AI impact and public sector savings

https://www2.deloitte.com/content/dam/insights/us/articles/3832_AI-augmented-government/DUP_AI-augmented-government.pdf

https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/10/ai-in-public-sector.pdf

Examples – Identifying children at risk

Predictive analytics

18 Using predictive analytics in the service of vulnerable citizens

SAS Case study:
Anomaly detection to detect changes in school attendance, for example—can flag potential higher risk.

Associative analysis can show who is linked to the child through records and addresses, and reveal their influence on upbringing.

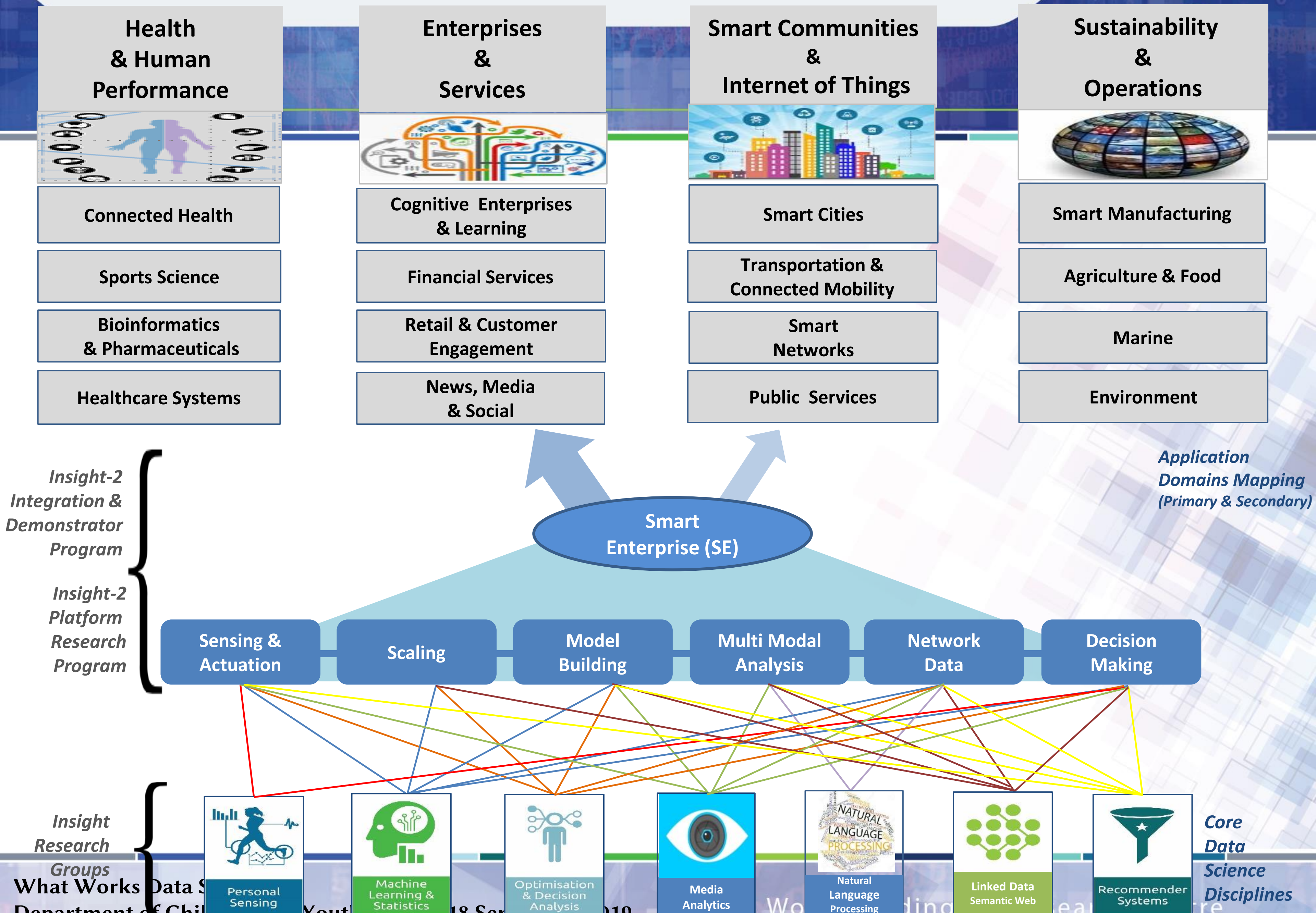
Predictive modelling based on historical events and correlation of data, can forecast outcomes, often with alarming accuracy.

For example, a U.S. county reported about 50 deaths per year of children in the system and 225 tragic outcomes over a five-year period. After developing an analytical model based on available data, which was applied to the cases, the model identified 176 of the 225 cases three to six months before the actual event.

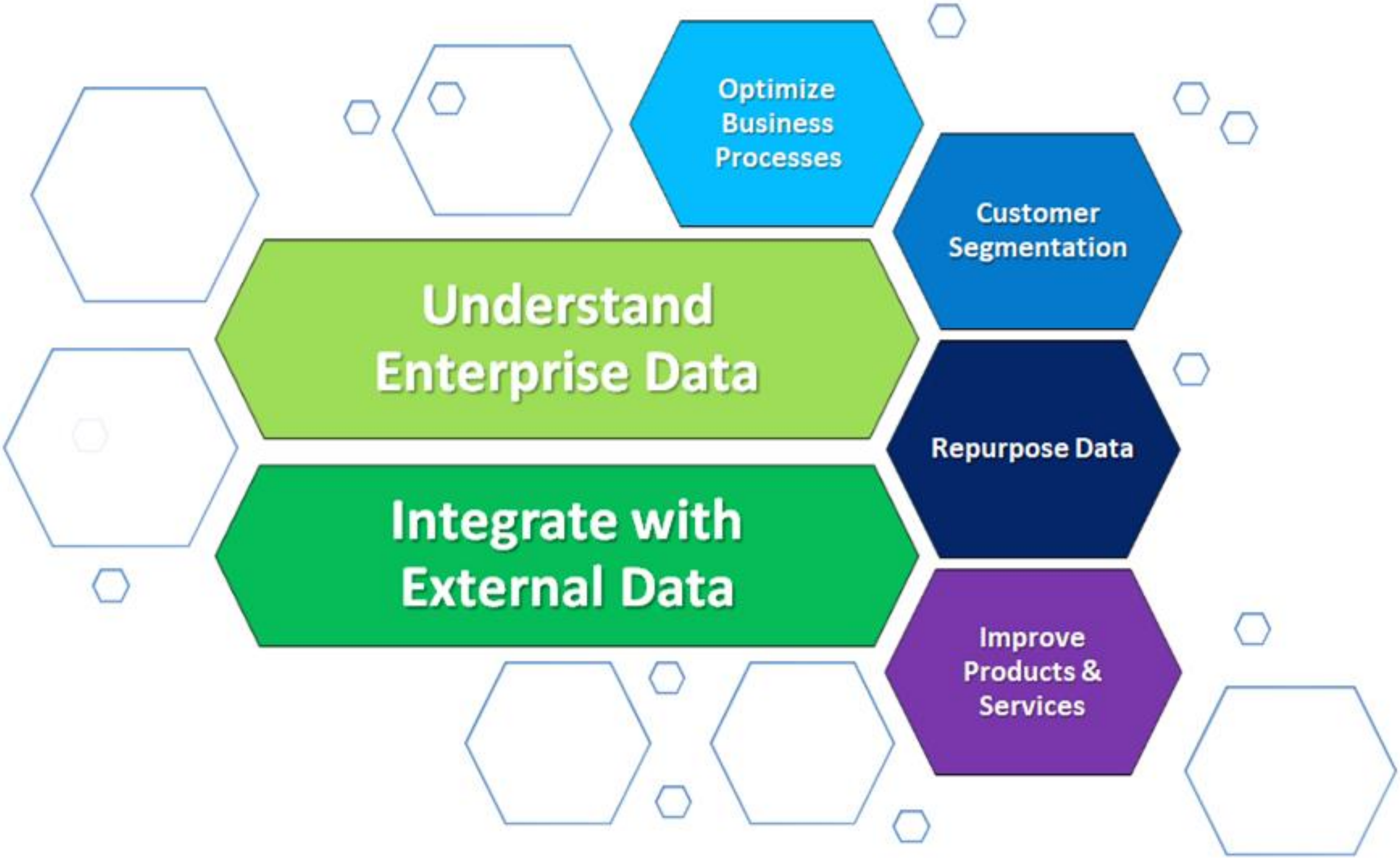
https://www.sas.com/en_ca/insights/articles/analytics/local/using-advanced-analytics-to-protect-at-risk-children.html

Example – Analytics in protecting children

https://www.youtube.com/watch?v=s_YzGaceyi8



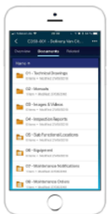
What Works Data S
Department of Chil Yout 18 Sep 2019
The Royal College of Physicians of Ireland, Dublin


Optimize Business Processes
Customer Segmentation
Understand Enterprise Data
Repurpose Data
Integrate with External Data
Improve Products & Services

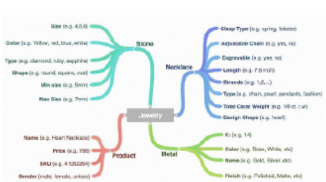
Insight
Centre for Data Analytics

Enterprise & Open Data

SharePoint



YouTube
Linked in



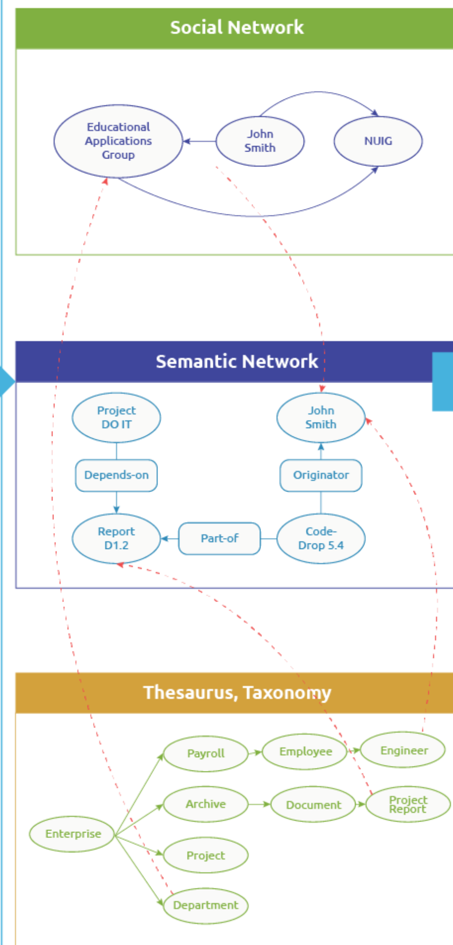
DPIA

GDPR Compliance

Data Analysis Components

Term Extraction	Taxonomy Extraction
Suggestion Mining	Emotion Detection
Sentiment Analysis	Machine Translation
Entity Linking	...
...	Link Completion
Image Classification	Salience Detection
Crowd Counting	...
...	...
Recommender System	...
...	Decision Support

Enterprise Knowledge Graph



Use Cases

Fidelity INVESTMENTS

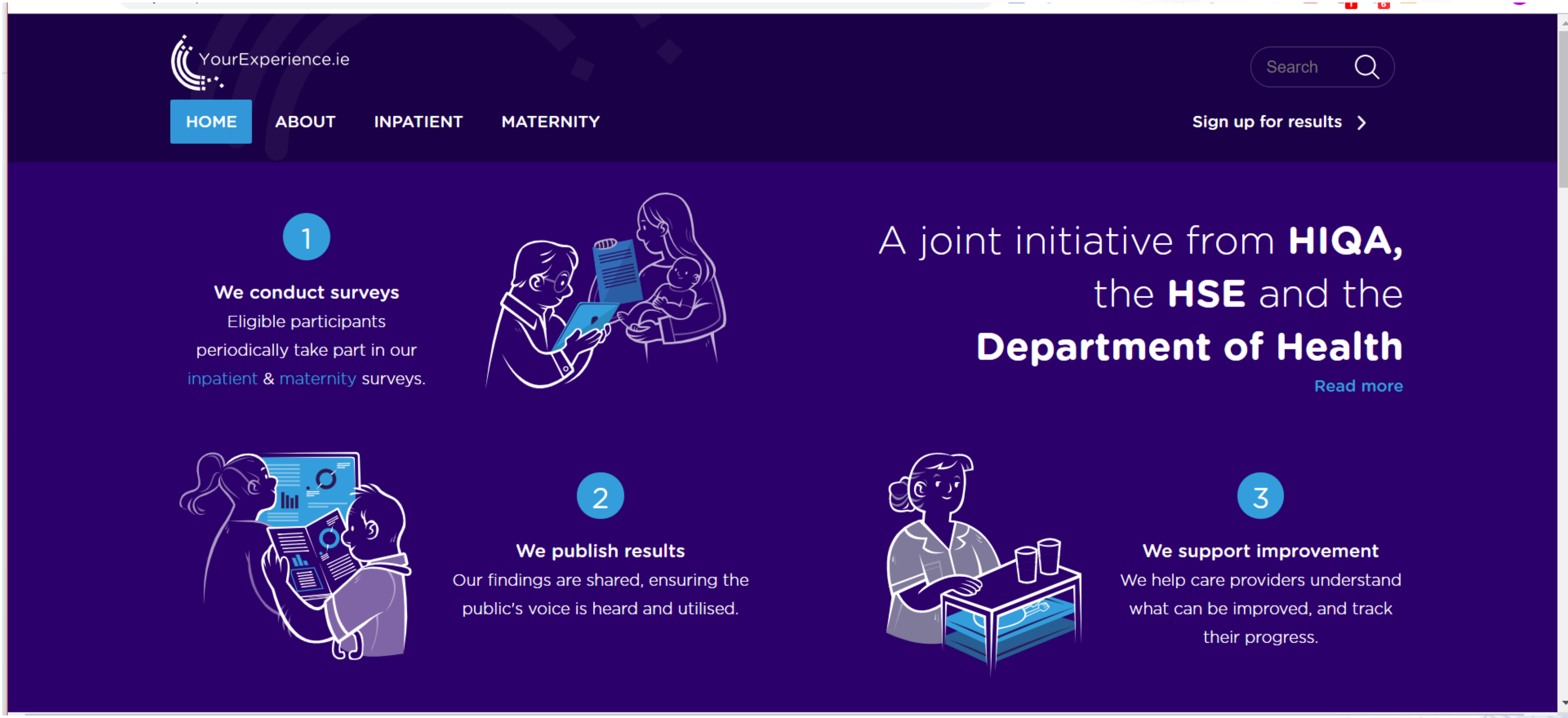
RTÉ

THE IRISH TIMES

Health Information and Quality Authority
An tÚdarás Um Fhaisnéis agus Cáilíocht Sláinte

Munich RE

Exemplar Insight Project with HIQA

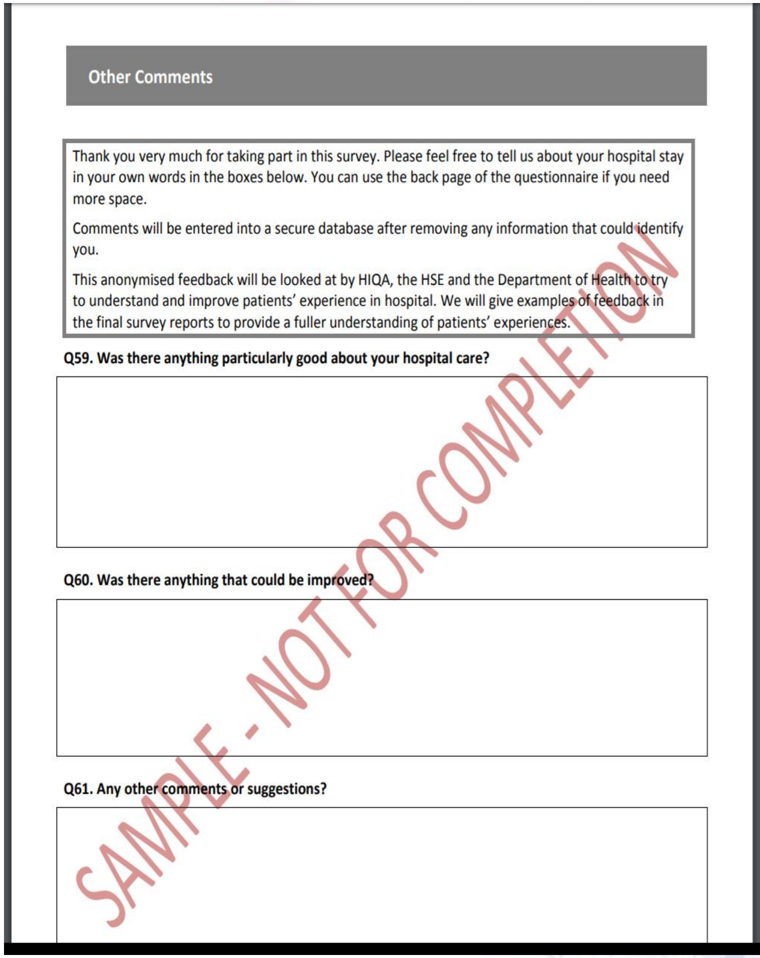


https://yourexperience.ie

Project Goal

Generating *actionable insights* from the free text responses of respondents to the 2017 & 2018 National Patient Experiences Surveys using *Qualitative and Computational Text Analytics methods.*

- What do patients like at the different stages of care?
- What are the recurring patterns of negative experience at the different stages of care and in what context do they occur?
- What suggestions were made by patients towards improvement?
- Are there significant disparity in negative experience across socio-demographic groups?


Other Comments

Thank you very much for taking part in this survey. Please feel free to tell us about your hospital stay in your own words in the boxes below. You can use the back page of the questionnaire if you need more space.

Comments will be entered into a secure database after removing any information that could identify you.

This anonymised feedback will be looked at by HIQA, the HSE and the Department of Health to try to understand and improve patients' experience in hospital. We will give examples of feedback in the final survey reports to provide a fuller understanding of patients' experiences.

Q59. Was there anything particularly good about your hospital care?

Q60. Was there anything that could be improved?

Q61. Any other comments or suggestions?

SAMPLE - NOT FOR COMPLETION

Analytical Framework

Stages of Care

1. Admission/Hospitalization
2. Patient care
3. Patient Treatment
4. Discharge
5. Other

https://drive.google.com/file/d/19pgfiy1UG0DEQL9T67zaZP4EFEM9DbvA/view

Activities

1- Admission

2-1-Care on the Ward
2-1-1- Patient Care on the Ward
2-1-2- Communication/Information Exchange with Patient
2-1-3- Psychological patient support
2-1-4- Relatives-related Care (Communication/Information Exchange)
2-1-5- Staff Management
2-1-6- Cleaning
2-1-7- Meal and Catering
2-1-8- Providing facilities

2-2- Care in Emergency
2-2-1- Patient Care in Emergency
2-2-2- Communication/Information Exchange with Patient
2-2-3- Psychological patient support
2-2-4- Relatives-related Care (Communication/Information Exchange)
2-2-5- Staff Management
2-2-6- Cleaning
2-2-7- Meal and Catering
2-2-8- Providing facilities

3-1- Treatment
3-1-1- Patient Treatment
3-1-2- Surgery/Procedures
3-1-3- Diagnosis

3-2- Operation briefing
3-3- Communication/Information Exchange with Patient
3-4- Communication/Information Exchange between Health Professionals

4-1- Discharge
4-2- Transfer
4-3- Discharge Communication
4-4- Payment

5-1- Parking

GDPR and Data Protection Training

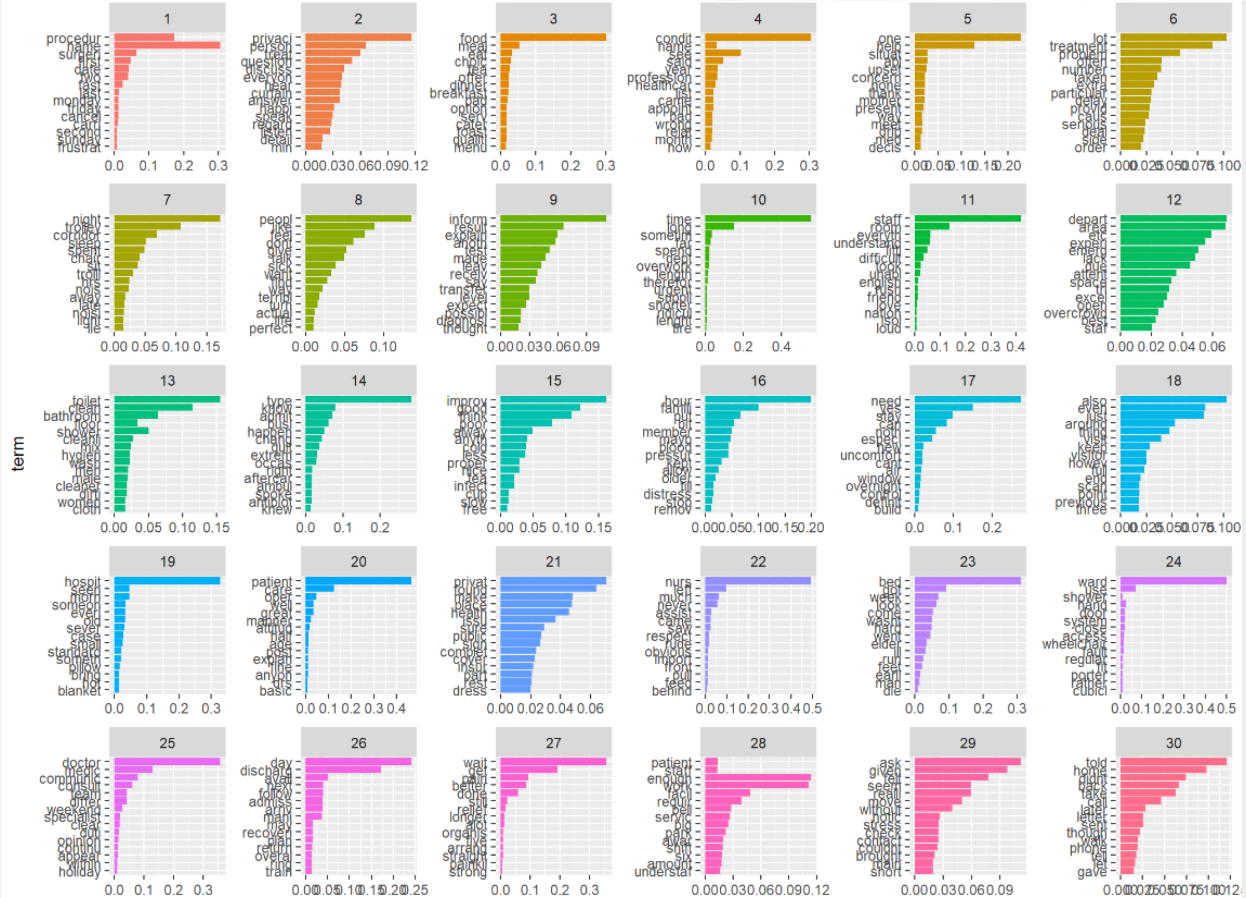
- Data protection training of the project members
- Held at the Insight Centre for Data Analytics in January
- Lasted for 1 hour and 30 minutes
- Ensured that project team members are fully aware of GDPR regulation and understand their responsibilities regarding compliance with respect to the NPES dataset

Exploratory Analysis – Topic Modelling

An algorithm that scans a set of documents (comments in our case) and examines how words and phrases co-occur in them, and automatically "learns" groups or clusters of words that best characterize those documents (or comments).

These sets of words represent a theme or topic.

http://rpubs.com/gboyegaojo/topics-neg-30-15-2-model

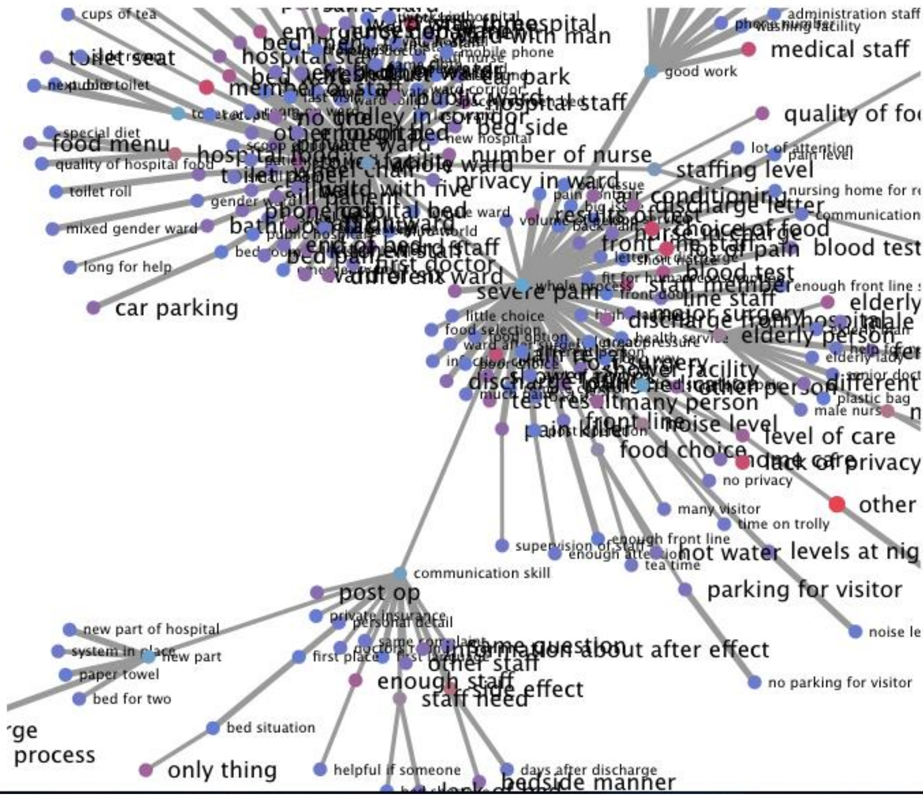


Topic hierarchies extracted from comments

Saffron
Knowledge Extraction Framework

Search ...

Taxonomy



Top Topics

- nursing staff
- length of time
- other patient
- waiting time
- private health insurance
- lack of communication
- communication between doctor
- family member
- choice of food
- cup of tea
- lot of pain
- member of staff
- health insurance
- lack of privacy
- stay in hospital

Excerpts of Results - ARC Pattern Mining

1. Export ARC-based annotations from database
2. Transform annotations into transactions
3. Explore transactions and generate patterns (aka rules)

Excerpts of Results – Recurring Patterns in comments

C:\Users\adegboyega.ojo\Documents\R works\NPES\m3-activity-290419-sorted-lift-s003-c5-l3.txt - Sublime Text (UNREGISTERED)

File Edit Selection Find View Goto Tools Project Preferences Help

Maintaining E-Voting Integrity through Blockchain | https://www.youtube.com/watch?v=h85PyEOxcss | extend.r | notes from discussion with segun | ---------------------------- | m3-activity-290419-sorted-lift-s003-c5-l3.txt

```
     lhs                                                    rhs                                                     support confidence      lift count
[1]  {A_Meal and Catering (Ward),
      R_Catering staff}                                  => {C_Catering staff insufficient procedures and practices} 0.005255954  0.6357616 98.408351    96
[2]  {A_Cleaning (Ward),
      C_Cleaning staff insufficient procedures and practices} => {R_Agency cleaners}                                  0.003449220  0.9130435 91.630435    63
[3]  {A_Meal and Catering (Ward),
      C_Catering staff insufficient procedures and practices} => {R_Catering staff}                                   0.005255954  0.9600000 80.065753    96
[4]  {A_Patient Care in Emergency,
      R_Trolley}                                         => {C_Patient left on a trolley}                             0.003284971  0.7500000 45.662500    60
[5]  {C_Long waiting time,
      R_Trolley}                                         => {C_Patient left on a trolley}                             0.003339721  0.6161616 37.513973    61
[6]  {C_Long waiting time,
      C_Patient left on a trolley}                       => {R_Trolley}                                               0.003339721  0.5545455 37.375545    61
[7]  {A_Patient Care in Emergency,
      C_Patient left on a trolley}                       => {R_Trolley}                                               0.003284971  0.5309735 35.786827    60
[8]  {A_Patient Care on the Ward,
      R_Trolley}                                         => {C_Patient left on a trolley}                             0.003011224  0.5670103 34.521478    55
[9]  {A_Discharge,
      C_Unsatisfactory discharge procedures}             => {R_Discharge protocol and arrangement}                    0.004489461  0.5616438 33.198785    82
[10] {C_Long waiting time,
      R_Discharge protocol and arrangement}              => {A_Discharge}                                             0.003065973  0.9032258 31.604252    56
[11] {C_Unsatisfactory discharge procedures,
      R_Discharge protocol and arrangement}              => {A_Discharge}                                             0.004489461  0.8631579 30.202259    82
[12] {C_Monitoring the cleaning standards,
      R_Toilet area}                                     => {A_Cleaning (Ward)}                                       0.005255954  0.9320388 23.448608    96
[13] {C_Cleaning staff insufficient procedures and practices,
      R_Agency cleaners}                                 => {A_Cleaning (Ward)}                                       0.003449220  0.8750000 22.013602    63
[14] {A_Patient Care on the Ward,
      C_Patient in mixed sex ward}                       => {R_Ward room}                                             0.003668218  0.5630252 21.880118    67
[15] {A_Staff Management (Ward),
      C_Night time}                                      => {C_Understaffed}                                          0.003120723  0.8382353 15.264574    57
[16] {C_Unfavourable Condition in ward,
      R_Toilet area}                                     => {A_Cleaning (Ward)}                                       0.004270463  0.5777778 14.535966    78
[17] {A_Staff Management (Ward),
      R_Nurse staff}                                     => {C_Understaffed}                                          0.005310704  0.7461538 13.587737    97
[18] {A_Meal and Catering (Ward),
      C_Food bad quality}                                => {R_Hospital food}                                         0.022611552  0.9740566 12.320737   413
[19] {A_Meal and Catering (Ward),
      C_Cold food}                                       => {R_Hospital food}                                         0.005365453  0.9702970 12.273182    98
[20] {C_Food bad quality,
      C_Limited menu}                                    => {R_Hospital food}                                         0.003339721  0.9682540 12.247340    61
[21] {A_Meal and Catering (Ward),
      C_Food bad quality,
      C_Limited menu}                                    => {R_Hospital food}                                         0.003339721  0.9682540 12.247340    61
[22] {A_Staff Management (Ward),
      R_Staff}                                           => {C_Understaffed}                                          0.019983575  0.6576577 11.976189   365
[23] {A_Staff Management (Ward),
      R_Medical doctor,
      R_Nurse}                                           => {C_Understaffed}                                          0.003832466  0.6542056 11.913325    70
[24] {C_Staff overworked,
      C_Understaffed,
      R_Staff}                                           => {A_Staff Management (Ward)}                               0.003011224  0.9166667 11.891276    55
[25] {C_Nurse overworked,
      C_Understaffed,
      R_Nurse}                                           => {A_Staff Management (Ward)}                               0.004434711  0.9101124 11.806252    81
[26] {A_Staff Management (Ward),
      R_Nurse}                                           => {C_Understaffed}                                          0.019271831  0.6365280 11.591410   352
```

Deep learning for annotating comments

```
 - 13s - loss: 13.0343 - crf_viterbi_accuracy: 0.8937 - val_loss: 13.1492 - val_crf_viterbi_accuracy: 0.6477
Epoch 111/120
 - 13s - loss: 13.0338 - crf_viterbi_accuracy: 0.8949 - val_loss: 13.1602 - val_crf_viterbi_accuracy: 0.6932
Epoch 112/120
 - 13s - loss: 13.0330 - crf_viterbi_accuracy: 0.8955 - val_loss: 13.1476 - val_crf_viterbi_accuracy: 0.6636
Epoch 113/120
 - 13s - loss: 13.0325 - crf_viterbi_accuracy: 0.8975 - val_loss: 13.1504 - val_crf_viterbi_accuracy: 0.6977
Epoch 114/120
 - 13s - loss: 13.0317 - crf_viterbi_accuracy: 0.8973 - val_loss: 13.1557 - val_crf_viterbi_accuracy: 0.6498
Epoch 115/120
 - 13s - loss: 13.0302 - crf_viterbi_accuracy: 0.8999 - val_loss: 13.1492 - val_crf_viterbi_accuracy: 0.6781
Epoch 116/120
 - 13s - loss: 13.0304 - crf_viterbi_accuracy: 0.8989 - val_loss: 13.1642 - val_crf_viterbi_accuracy: 0.6446
Epoch 117/120
 - 13s - loss: 13.0294 - crf_viterbi_accuracy: 0.9002 - val_loss: 13.1577 - val_crf_viterbi_accuracy: 0.6552
Epoch 118/120
 - 13s - loss: 13.0287 - crf_viterbi_accuracy: 0.9016 - val_loss: 13.1863 - val_crf_viterbi_accuracy: 0.6757
Epoch 119/120
 - 13s - loss: 13.0279 - crf_viterbi_accuracy: 0.9020 - val_loss: 13.1827 - val_crf_viterbi_accuracy: 0.6930
Epoch 120/120
 - 13s - loss: 13.0271 - crf_viterbi_accuracy: 0.9012 - val_loss: 13.1576 - val_crf_viterbi_accuracy: 0.6808
              precision    recall  f1-score   support

          BA       0.40      0.36      0.38      9139
          BC       0.19      0.15      0.17      6500
          BR       0.65      0.56      0.60      5641
          IA       0.30      0.28      0.29     17131
          IC       0.26      0.22      0.24     27298
          IR       0.36      0.24      0.29      1802
           O       0.79      0.83      0.81    165874
         PAD       1.00      1.00      1.00    261495

    accuracy                           0.85    494880
   macro avg       0.50      0.46      0.47    494880
weighted avg       0.84      0.85      0.84    494880
```

If we succeed in our machine annotation experiments, we may be able to reduce subsequent comment processing time by as much as 80%

Challenges

- No explicit strategy or thinking about how to harness data and analytics-driven improvement.
- Limited data and analytics capabilities in government agencies and departments
- Quality of data available for analytics is often low, significant investment is required in refining the raw data into form that is fit for analytics
- Level of trust between citizens and government is low when it comes to sharing of information with government.

Any question?

adegboyega.ojo@nuigalway.ie